



Gargon, E., Williamson, P. R., Blazeby, J. M., & Kirkham, J. J. (2019). Improvement was needed in the standards of development for cancer core outcome sets. *Journal of Clinical Epidemiology*, 112, 36-44.  
<https://doi.org/10.1016/j.jclinepi.2019.04.006>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY-NC-ND

Link to published version (if available):  
[10.1016/j.jclinepi.2019.04.006](https://doi.org/10.1016/j.jclinepi.2019.04.006)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Elsevier at <https://www.sciencedirect.com/science/article/pii/S0895435618310795> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

ORIGINAL ARTICLE

# Improvement was needed in the standards of development for cancer core outcome sets

Elizabeth Gargon<sup>a,\*</sup>, Paula R. Williamson<sup>a</sup>, Jane M. Blazeby<sup>b</sup>, Jamie J. Kirkham<sup>a</sup>

<sup>a</sup>Department of Biostatistics, MRC North West Hub for Trials Methodology Research, University of Liverpool, Liverpool, United Kingdom

<sup>b</sup>MRC ConDuCT II Hub for Trials Methodology Research and National Institute for Health Research Bristol Biomedical Research Centre, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

Accepted 9 April 2019; Published online 19 April 2019

## Abstract

**Objective:** The Core Outcome Set—STAndards for Development (COS-STAD) contains 11 standards (12 criteria) that are deemed to be the minimum design recommendations for all core outcome set (COS) development projects. Cancer is currently the disease area with the highest number of published COSs and is a major cause of worldwide morbidity and mortality. The aim of this study was to provide a baseline of cancer COS standards.

**Study Design and Setting:** Systematic reviews of COSs have identified 307 published COS studies. Cancer COSs were eligible for inclusion. Two reviewers independently assessed each of the COSs against the 12 criteria.

**Results:** Forty-nine cancer COSs were included; none met all 12 criteria representing the 11 minimum standards assessed in this study (range = 4–11 criteria, median = 6 criteria). All studies met the four scope standards, eight (16%) met all three standards for stakeholders involved, and two (4%) met all four standards for consensus process standards.

**Conclusion:** With the exception of “scope” specification, there is much need for improvement. Poor reporting often made it challenging to assess whether minimum standards were met. The consensus process criteria were most difficult to assess, particularly those that required an assessment of being a priori. This is the first application of COS-STAD criteria to studies that have developed COSs and provides a baseline of cancer COS standards of development. © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Core outcome set; Minimum standards; Cancer; Research methodology; COS; Study design

Conflict of interest: E.G., J.M.B., and P.R.W. are members of the COMET Management Group and co-applicants on grants to support COMET and related work. J.M.B., J.J.K., and P.R.W. were involved in the development of COS-STAD. J.M.B., P.R.W., and J.J.K. have been involved in the development of COS.

**Funding:** This work was supported by the NIHR Senior Investigator Award (NF-SI\_0513-10025). Professor Paula Williamson is a senior investigator of the National Institute for Health Research (NIHR). The views expressed in this article are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. J.M.B. is supported by the MRC ConDuCT-II Hub and the NIHR Biomedical Research Center at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. J.M.B. is also a senior investigator of the NIHR.

\* Corresponding author. Tel.: +440 151 794 9727.

E-mail address: [e.gargon@liverpool.ac.uk](mailto:e.gargon@liverpool.ac.uk) (E. Gargon).

## 1. Introduction

To make well-informed decisions about health care, we need to be able to compare and contrast research findings on the basis of the same outcomes. Core outcome sets (COSs) represent the minimum important outcomes that should be measured and reported in all research studies for a specific condition. The use of COSs will improve the quality of evidence used in health care decision-making, ultimately translating to improved health care for patients. For COSs to be successfully implemented, they need to be easily accessible to researchers and other key groups, developed using rigorous methods, and reported clearly. The Core Outcome Measures in Effectiveness Trials (COMET) Initiative was set up to help achieve this. The development and continued maintenance of the COMET database [1,2], through rigorous systematic reviews [3–7], means that COSs are now easily accessible to users of COSs. The COMET Handbook advocates the use of rigorous methods through an accumulation of current knowledge in the area of COS development and will be updated

**What is new?****Key findings**

- No core outcome set (COS) met all of the 12 criteria representing the 11 minimum standards assessed in this study (range = 4–11 criteria, median = 6 criteria).
- Poor reporting often made it challenging to assess whether the minimum standards had been met.

**What this adds to what was known?**

- This is the first application of Core Outcome Set—STAndards for Development (COS-STAD) criteria to studies that have developed COSs and provides a baseline of cancer COS standards of development.
- COS-STAD was not published until 2017; all included COS studies were carried out before this publication date. Therefore, this assessment is not a criticism of these studies or the study authors, but rather a baseline against which future comparisons of cancer COSs can be made.

**What is the implication and what should change now?**

- This current review provides guidance on how to compare a published COS to the standards (Table 2). This will further facilitate users to assess whether a COS has been well developed.
- This study identified the need to separate standard number 9 into two criteria, considering scoring process and consensus definition separately. We recommend this separation for future users of COS-STAD.

periodically to continue to inform good methodological practice in this area [8]. Furthermore, the Core Outcome Set—STAndards for Development (COS-STAD) recommendations have been established to improve the methodological approach for developing COSs and help users to assess whether a particular COS has been developed using a reasonable approach [9].

Research in the area of COS development is becoming more prevalent, but it is still quite new; we would therefore expect improvements in methodological standards in the coming years. To be able to assess this, however, a baseline, against which we can compare future quality of COS development and measure improvements in methodological standards, is necessary. COSs are developed across a wide range of disease areas. Cancer is currently the disease area with the highest number of published COSs. Furthermore, cancer is a major cause of worldwide morbidity and mortality, has

substantial variability in populations and treatments, and covers a wide range of diverse clinical areas. Treatments aim to cure the disease but are associated with multiple side effects; trials are commonly performed, and major clinical oncology organizations exist worldwide. It is essential to ensure that outcomes in cancer trials reflect issues of importance to patients and health care professionals through the use of well-developed COSs. The aim of this study was therefore to provide a baseline of cancer COS standards of development against which future comparisons can be made.

**2. Methods****2.1. Identification of COS studies**

As of the end of 2017, a regularly updated systematic review of COSs [3–7] had identified 307 published COSs for research studies covering a wide range of different health areas. Only COSs developed for cancer were eligible for inclusion in this evaluation.

**2.2. Pilot study**

COS-STAD contains 11 standards that are deemed to be the minimum design recommendations for all COS development projects [9]. The recommendations focus on three key domains: scope, stakeholders, and consensus process.

A pilot study was carried out to inform the process of assessment used in this study. One COS was randomly selected. Three of the authors (E.G., P.R.W., and J.J.K.) independently read the COS and made an assessment against each of the 11 criteria listed in Table 1. Assessments were compared, and a discussion occurred between the three authors to define how the process should be applied. This included the criteria used for assessment and the sources of information that would be used to identify supporting information. Following this, a further five randomly selected COSs were independently assessed by two of the authors (E.G. and J.J.K.), and the process was further refined. Of note, it was agreed that criteria number 9 (“A scoring process and consensus definition is described a priori”) would be split into two for assessment purposes. It became apparent that in the five COSs assessed in the pilot study that a study team could describe an a priori definition for one of these only (e.g., provide a description of a scoring process but not provide a consensus definition or vice versa), resulting in different assessments for each part of this criterion and should therefore be assessed separately. Hereafter, they will be referred to as 9a (scoring process) and 9b (consensus definition). A total of 12 criteria were therefore assessed in this study, representing the 11 minimum standards.

**2.3. Process**

Two reviewers (E.G. and J.J.K.) independently assessed each of the cancer COSs against the 12 criteria of

**Table 1.** Original COS-STAD criteria

Domain	Standard number	Standard
Scope specification	1	The research or practice setting(s) in which the COS is to be applied
	2	The health condition(s) covered by the COS
	3	The population(s) covered by the COS
	4	The intervention(s) covered by the COS
Stakeholders involved	5	Those who will use the COS in research
	6	Health care professionals with experience of patients with the condition
	7	Patients with the condition or their representatives
Consensus process	8	Initial list of outcomes considered both health care professionals' and patients' views
	9	A scoring process and consensus definition was described a priori
	10	Criteria for including/dropping/adding outcomes were described a priori
	11	Care was taken to avoid ambiguity of language used in the list of outcomes

Abbreviations: COS-STAD, Core Outcome Set–STAndards for Development.

development. Each criterion was assessed as yes (meeting that standard), no (not meeting that standard), or unsure (it was unclear whether the criteria had been met). Further details of the assessment are described in [Table 2](#).

Verbatim text was extracted to support the assessment being made and to aid discussion. Assessments were compared, and a third reviewer consulted where there was uncertainty. P.R.W. and J.M.B. were consulted for methodological queries, and J.M.B. (surgical oncologist), for clinical queries. Where the development process was described across multiple articles for an individual COS, a global assessment was made for each of the standards.

### 2.3.1. Sources of information

Articles describing the development of the COS were eligible for providing supporting information for each of the standards. This included main study publications and protocols. The COMET database was not used as a source of information in this study as it is a secondary source of information (populated by the systematic reviews used in this study to identify cancer COSs) and does not contain any additional details necessary for this study.

### 2.4. Data analysis and presentation of results

The median and range were used to summarize the 12 criteria (representing the 11 minimum standards) met across all of the included cancer COS studies. Percentage frequencies were used to report the number of studies that met each standard.

## 3. Results

Forty-nine cancer COSs were included. An overview of the minimum standard assessments is provided in [Table 3](#), and by study in [Appendix 1](#). No COS met all of the 12 criteria representing the 11 minimum standards assessed in this study (range = 4–11 criteria, median = 6 criteria).

### 3.1. Scope specification

All studies met all four standards for scope.

#### 3.1.1. Standard number 1: the research or practice setting(s) in which the COS is to be applied

All studies met this criterion: 40/49 (82%) stated that the COS was intended for use in clinical trials, 4/49 (8%) stated that the COS was intended for use in clinical research, two (4%) stated that it was intended for use in clinical trials and routine clinical practice, and three (6%) for research and routine clinical practice.

#### 3.1.2. Standard number 2: the health condition(s) covered by the COS

All studies met this criterion. Five COSs were developed for cancer (nonspecific), and four specifically for complications of the treatments of cancer. Nine COSs were developed for prostate cancer, five for cancer of the colon and/or rectum, and three for ovarian cancer. Two were developed for non-small cell lung cancer, two for Hodgkin's and non-Hodgkin's lymphomas, and two COSs were developed for acute myeloid leukemia. The remaining seventeen unique cancer COS areas are listed in [Appendix 2](#).

#### 3.1.3. Standard number 3: the population(s) covered by the COS

All studies met this criterion. In 10 of the cancer COSs, it was not explicitly stated that the COS was intended for an adult population only, but this was assumed because the cancer covered by the COS was an adult-only cancer (e.g., prostate cancer). Further assumptions were made for an additional thirteen COSs that related to cancers that may occur in children and adults but did not make any statement about the population that the COS was intended. These 13 COSs were deemed to be implicitly for adults only because they did not refer to children specifically.

**Table 2.** Assessment criteria: guidance on how to compare a published COS to the standards

Domain	Standard number	Standard	Features to be considered	Standard met	Unclear whether standard is met	Standard not met
Scope specification	1	The research or practice setting(s) in which the COS is to be applied should be considered	Descriptions include (but not limited to) health research, specific study types (e.g., trials, observational studies, and longitudinal studies), routine care, audit, and registries. The developed COS might apply to a single setting or a combination of settings.	Evidence that the setting(s) (or context of use) in which the COS is to be applied has been considered.	The setting(s) (or context of use) in which the COS is to be applied is unclear or is not described.	An explicit statement that this was not considered.
	2	The health condition(s) covered by the COS should be considered	This could be general (e.g., cancer) or specific (e.g., lung cancer).	Evidence that the health condition(s) has been considered.	The health condition(s) is unclear or is not described.	An explicit statement that this was not considered.
	3	The population(s) covered by the COS should be considered	Characteristics should be relevant to the population. The following were considered for cancer COSs: 1 Disease stage 2 Age of population	Evidence that the population has been considered.	Population characteristics are discussed (e.g., in the introduction or discussion of the article) but not in the context of the COS being developed. Population is unclear or not described.	An explicit statement that this was not considered.
	4	The intervention(s) covered by the COS should be considered	1 This could be all/any intervention or specific intervention types (e.g., drug, surgery)	Evidence that the intervention has been considered.	Interventions are discussed (e.g., in the introduction or discussion of the article) but not in the context of the COS being developed. Intervention is unclear or not described.	An explicit statement that this was not considered.
Stakeholders involved	5	Those who will use the COS in research	Descriptions include (but not limited to) the following: 1 Clinical trialists (including trial investigators, members of trial groups) 2 Researchers (including methodologists, experts with expertise in the conduct of studies, scientists) 3 Industry representatives	COS article includes description of i. <b>OR</b> ii. <b>OR</b> iii.	It was not clear whether those who will use the COS in research were included.	It was clear that those who will use the COS in research were not included.
	6	Health care professionals with experience of patients with the condition	Health care professionals (including all disciplines such as oncologists, urologists, pharmacists, surgeons, nurses, etc.) <sup>b</sup>	COS article includes description of health care professionals.	It was not clear whether health care professionals with experience of patients with the condition were included.	It was clear that health care professionals with experience of patients were not included.

(Continued)

Table 2. Continued

Domain	Standard number	Standard	Features to be considered	Standard met	Unclear whether standard is met	Standard not met
	7	Patients with the condition or their representatives	Descriptions include (but not limited to) the following: 1 Patients with the condition 2 Patient representatives 3 Patient advocates 4 Parents 5 Carers	COS article includes description of i. <b>OR</b> ii. <b>OR</b> iii. <b>OR</b> iv. <b>OR</b> v.	It was not clear whether patients/patient representatives were included.	It was clear that patients/patient representatives were not included.
Consensus process	8	Initial list of outcomes considered both health care professionals' and patients' views	1 Health care professionals (see <a href="#">standard number 6 descriptors</a> ) 2 Patient (see <a href="#">standard number 7 descriptors</a> )	Initial list clearly included i. <b>AND</b> ii.	It was not clear whose views' were considered.	The initial list only included i. <b>OR</b> ii. The initial list did not include either view.
	9a	A scoring process was described a priori <sup>a</sup>	Any description of a scoring system to rate outcomes [8], including (but not limited to) the following: 1 A Likert scale or similar 2 Ranking of outcomes 3 Allocation of points	It does describe a scoring process, and it is clear that this <b>was</b> a priori <sup>a</sup>	It does describe a scoring process but does <b>not</b> state whether this was a priori (or when it was described/defined). It does <b>not</b> mention a scoring process.	It clearly has not used a scoring process. It does mention a scoring process, but states this <b>was not</b> a priori, or text about methods confirms it.
	9b	A consensus definition was described a priori <sup>a</sup>	There are numerous ways to define the consensus criteria, commonly these relate to a mean or median value for each outcome or a percentage of participants scoring an outcome as important [8].	It does provide a definition of consensus, and it is clear that this <b>was</b> a priori <sup>a</sup> .	It does mention consensus (or a synonym) or provides a description of definition, but it is <b>not</b> clear whether this was a priori/when it was described/defined. It does <b>not</b> mention consensus.	It clearly has not used consensus methods. It does mention consensus (or a synonym), does provide further description of definition, but states this <b>was not</b> a priori, or text about methods confirms it.
	10	Criteria for including/dropping/adding outcomes were described a priori <sup>a</sup>	A description of this process (e.g., a description of including AND dropping AND adding outcomes).	It does include a description of including AND dropping AND adding outcomes, as well as stating this <b>was</b> a priori <sup>a</sup> .	It does include a description of including AND dropping AND adding outcomes but <b>does not</b> state whether this was a priori/when it was described/defined. It does <b>not</b> include any description about including/adding/dropping outcomes. It does <b>not</b> describe <b>all</b> three elements of this criterion (including AND dropping AND adding outcomes).	It does include a description of including/dropping/adding outcomes, but states this <b>was not</b> a priori, or text about methods confirms it.

(Continued)



Table 2. Continued

Domain	Standard number	Standard	Features to be considered	Standard met	Unclear whether standard is met	Standard not met
	11	Care was taken to avoid ambiguity of language used in the list of outcomes	Consideration for language should be taken into account when describing outcomes to different stakeholder groups. An example might be the use of both plain language descriptions and medical terms, with these tested in a pilot study for understanding [9].	A clear description of methods/steps taken to avoid ambiguity of language.	Some suggestion that this may have been done, but not clearly described. No evidence of consideration given to ambiguity of language.	An explicit statement that this was not considered.

Abbreviations: COS-STAD, Core Outcome Set–STAndards for Development.

<sup>a</sup> “a priori” as assessed by inclusion in a protocol or when stated “a priori” in the study report without a protocol to verify this (we have taken this at face value).

<sup>b</sup> Clinicians, physicians, clinical investigators, and medical faculty were all different descriptors used for this in the cancer COS articles.

### 3.1.4. Standard number 4: the intervention(s) covered by the COS

All studies met this criterion. Where the authors referred to “evaluating treatments,” “cancer treatments,” or a COS for “all trials,” we took this to mean all interventions. Twenty-six (53%) COSs were developed for all/any interventions, eight (16%) for drug interventions, four (8%) for procedures, three (6%) for surgical interventions, two (4%) for screening, and the remaining six (12%) were for other specific interventions (preoperative treatment strategies, adjuvant treatment, larynx preservation strategies, vaccination, exercise, and compression interventions).

### 3.2. Stakeholders involved

Eight (16%) studies met all three standards for stakeholders involved.

#### 3.2.1. Standard number 5: those who will use the COS in research

Thirty-six studies (74%) met this standard. Assumptions were made based on the author affiliations (where it was clear that the authors contributed to the COS development process) or from the participant list affiliations when these were provided, for 10 COSs. These 10 did not explicitly

Table 3. Cancer COS minimum standard assessments (*N* = 49)

Domain	Standard number	Standard	Standard met, <i>n</i> (%)	Standard unclear, <i>n</i> (%)	Standard not met, <i>n</i> (%)
Scope specification	1	The research or practice setting(s) in which the COS is to be applied	49 (100)	0	0
	2	The health condition(s) covered by the COS	49 (100)	0	0
	3	The population(s) covered by the COS	49 (100)	0	0
	4	The intervention(s) covered by the COS	49 (100)	0	0
Stakeholders involved	5	Those who will use the COS in research	36 (74%)	7 (14)	6 (12)
	6	Health care professionals with experience of patients with the condition	35 (71)	8 (16)	6 (12)
	7	Patients with the condition or their representatives	13 (27)	1 (2)	35 (71)
Consensus process	8	Initial list of outcomes considered both health care professionals' and patients' views	8 (16)	5 (10)	36 (74)
	9a	A scoring process was described a priori	5 (10)	44 (90)	0
	9b	A consensus definition was described a priori	6 (12)	43 (88)	0
	10	Criteria for including/dropping/adding outcomes were described a priori	2 (4)	47 (96)	0
	11	Care was taken to avoid ambiguity of language used in the list of outcomes	5 (10)	44 (90)	0

Abbreviations: COS-STAD, Core Outcome Set–STAndards for Development.

state that researchers were included as participants, but we assumed their inclusion when research institutions were listed as affiliations.

Seven studies (14%) did not state that those who will use the COS in research were involved in the COS development and there was not enough evidence to support an assumption being made; therefore, these were categorized as unclear. Examples include where it was not clear that the authors were the panel/participants involved (and no other statement was made in relation to stakeholders), participants were described as conference participants, but their backgrounds were not described or they were described as “experts” without any further description.

Six studies (12%) did not meet this standard. In these instances, the stakeholders involved were clearly described and did not include those who will use the COS in research.

### 3.2.2. Standard number 6: health care professionals with experience of patients with the condition

Thirty-five studies (71%) clearly met this standard. Assumptions were made for eight COSs based on the author affiliations (where it was clear that the authors are the group who developed the COS) or from the participant list affiliations when these were provided. These eight did not explicitly state that health care professionals (HCPs) were included as participants, but we assumed their inclusion when clinical care settings were listed as affiliations.

Eight (16%) did not state explicitly that HCPs were involved in the COS development and there was ambiguity in the description provided; these were, therefore, categorized as unclear. Examples echoed those given for standard number 5.

Six studies (12%) did not meet this standard. In these instances, the stakeholders involved were clearly described and did not include HCPs in the process.

### 3.2.3. Standard number 7: patients with the condition or their representatives

Thirteen studies (27%) met this standard. Eight studies described including patient advocates or representatives, and five included patients themselves. The number of patients in six of these studies was not reported. The percentage of patients with the condition or their representatives in the remaining seven studies ranged from 9% to 68% in studies with mixed participants.

Thirty-five studies (71%) did not meet this standard. In these instances, the stakeholders involved were clearly described and did not include patients in the process. It was unclear in one study whether patients were included as participants were described as conference delegates without any further description.

## 3.3. Consensus process

Two studies (4%) met all four standards [five criteria] for the consensus process.

### 3.3.1. Standard number 8: the initial list of outcomes considered both health care professionals' and patients' views

Eight studies (16%) met this standard. Two of these five included patient-reported outcome (PRO) data (i.e., studies that have collected data from patients using a patient-reported outcome measure [PROM]) as well as trial data, and therefore on face value have included the patient perspective in the process.

Five studies (10%) did not clearly state whose views were considered when generating the initial list and so, therefore, have been categorized as unclear. One of these studies included PROM items in the review (PROM items only not PRO data). Patients may or may not have been involved in the PRO development, but it was beyond the scope of this study to research PROM development; we cannot therefore say if they considered patients' views.

Thirty-six studies (74%) did not meet this standard and did not consider both HCPs' *and* patients' views when generating the initial list of outcomes used in the COS development. These studies considered trial data, clinical trials literature, or clinical guidelines only (hence did not consider patients' views).

### 3.3.2. Standard number 9a: a scoring process was described a priori

Five studies (10%) met this standard. It was unclear whether the remaining 44 (90%) met this standard. Ten studies described a scoring process, but it was not clear whether this process was determined a priori; 28 studies did not describe specific methods relating to scoring, and six studies did not describe the methods used at all.

### 3.3.3. Standard number 9b: a consensus definition was described a priori

Six studies (12%) met this standard. It was unclear whether this standard was met for the remaining 43 (88%). In 26 studies, it was unclear whether the consensus definition was defined a priori; eleven studies did not describe a process of consensus and therefore did not provide a definition, and the remaining six studies did not describe their consensus methods at all.

### 3.3.4. Standard number 10: criteria for including/dropping/adding outcomes were described a priori

Two studies (4%) met this standard. It was unclear whether this standard was met for the remaining 47 studies (96%). The detail of all three elements of this standard was not clearly described in six studies, and for two studies that did describe this process, it was not possible to assess the a priori element of this criterion. Thirty-three studies had no description of this process, and six did not describe their methods overall, making it impossible to assess whether this criterion had been considered.



### 3.3.5. Standard number 11: care was taken to avoid ambiguity of language used in the list of outcomes

Five studies (10%) met this criterion. In one study, the outcomes were formulated with examples in parentheses for some PROs; and in another study, the patient representatives received a glossary of terms before completing the survey. Although this information is limited as COS-STAD is being applied retrospectively here, we have interpreted this as some consideration of the language used to describe outcomes. This criterion was clearly met in the other three studies, where the questionnaires were assessed, by patients, for face validity and comprehension before use.

For 44/49 studies (90%), there was no evidence that care was taken to avoid ambiguity of language, and therefore it was unclear whether they met this criterion.

## 4. Discussion

This is the first application of the COS-STAD criteria to studies that have developed COSs. Forty-nine cancer COSs were included. No COS met all of the minimum standards, with most studies meeting half of the standards. However, COS-STAD was not published until 2017, which was after all of these studies started. Therefore, this assessment is not a criticism of these studies or the study authors, but rather a baseline against which future comparisons can be made. Furthermore, this current review provides guidance on how to compare a published COS to the standards (Table 2). This will further facilitate users to assess whether a COS has been well developed.

Standards in the scope domain were well met. One explanation for this could be that these scope criteria are synonymous with the “PIC” part of the PICO format that is often used to formulate a research question [10]. For the purposes of this study, there was an assumption that a COS was developed for adults only unless stated otherwise. This was deemed to be clinically relevant for cancer COSs, and consideration should be given to other disease areas and whether the same assumption is correct. When assessing standards in the scope domain, we observed that multiple COSs were developed for some cancers, for example, nine COSs have been developed for prostate cancer alone. At times, this may reflect unnecessary duplication, but often they had relevant and appropriate differences in scope; for example, differences in stage or type such as advanced or localized cancer. COS-STAD can be used as a tool to help users of COSs to assess whether a COS has been well developed. Users will need to use their own judgment regarding the applicability of the COS (scope) for the purpose they require [9].

Most COS studies did include those who will use the COS in research and HCPs in the development process, whereas just over a quarter included patients or patient representatives in the process. This is reflective of COSs in general and is not a methodological problem specific to

cancer. Recent research has shown an improvement in patient participation in more recently published COSs [6,11]. Although thirteen studies did include patients in COS development, there is still great variability in the level of participation of patients in COS development. This also needs to be taken into account when deciding if a COS was well developed.

The consensus process standards were the most difficult to compare against, particularly those that required an assessment of being specified a priori. Only four studies stated that they had a protocol, of which two were published and one we were able to obtain from study authors. In the remaining studies, because we were applying this retrospectively, we took it at face value when the COS report stated that something was specified a priori because we were unable to verify this in a protocol. Although making the protocol publicly available for a COS development study was not agreed on as a COS-STAD minimum standard, it was suggested that the availability of a protocol would ensure that the methods are explicitly documented before the COS development project starts, thus promoting research integrity and transparency of the finalized COS. This will make it easier for users to assess the COS against COS-STAD for future studies.

This study identified the need to consider the scoring process and consensus definition separately. We recommend this separation for future users of COS-STAD. With regard to criteria 9a (scoring), an observation was that in those that did not meet this standard, the method of development used for consensus in all of these studies was some form of meeting or workshop to decide on the COS, but detailed description of what was done at those meetings was missing. Further consideration is needed about the applicability and suitability of this criterion for all methods that might be used for COS development.

Standard number 10, the criterion for including/dropping/adding outcomes, was only met by two studies. The a priori part of the assessment was unclear for a further two studies, but the majority did not mention anything about this process or did not describe it in sufficient detail to be able to assess whether this criterion was met. This suggests that perhaps COS developers do not fully understand the implications or importance of this aspect of the development process, and detail might be lacking as it might be considered too much information for publication. It should be noted that a lack of reporting of methods does not necessarily mean that these methodological aspects were not considered. This standard is also reflected in the Core Outcome Set-STANDards for Reporting (COS-STAR) as a reporting requirement [12]. However, COS-STAR was not published until 2016, which is after most of these included COS studies were published. It is therefore unrealistic to expect these reporting standards to be implemented in these studies, so it is to be hoped that evidence of consideration of this criterion (and indeed the other standards) will be included in study reports in the future.

COS-STAD focusses on the main design principles for COS development, whereas COS-STAR is exclusive to the reporting of COS studies. As already highlighted, issues with reporting were one of the main limitations of this study as it made it difficult, sometimes impossible, to assess consideration of the COS-STAD standards. We did obtain protocols when it was stated that a protocol was available, but this was very few. As such, we were unable to distinguish development plans from poor final reporting. Furthermore, this meant that assumptions were sometimes made to enable a judgment to be made. Another limitation was that the studies being assessed were published before COS-STAD and therefore could not have been informed by the development standards. Another potential limitation is that we identified cancer COSs from existing systematic reviews of COSs and did not conduct a separate search. To address this potential limitation, the list of COSs was reviewed by a relevant expert (J.M.B.) for completeness. The inclusion of cancer/COS methodology experts was a further strength of this study, as was the independent dual extraction and assessment of data.

This study aimed to provide a baseline of cancer COS standards of development against which future comparisons can be made. With the exception of scope specification, there is much need for improvement. Poor reporting made it challenging to assess whether the minimum standards had been met for all stakeholder involvement and consensus process standards. With the publication of methodological evidence [13,14] and guidance [8,9,15], as well as reporting standards [12], improvement is expected over time.

### CRedit authorship contribution statement

**Elizabeth Gargon:** Conceptualization, Methodology, Data curation, Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **Paula R. Williamson:** Conceptualization, Methodology, Formal analysis, Writing - review and editing. **Jane M. Blazeby:** Formal analysis, Writing - review & editing. **Jamie J. Kirkham:** Conceptualization, Methodology, Writing - review & editing.

### Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2019.04.006>.

### References

- [1] Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET Initiative database: progress and activities from 2011 to 2013. *Trials* 2014;15(1):279.
- [2] Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET initiative database: progress and activities update (2014). *Trials* 2015;16(1):515.
- [3] Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. *PLoS One* 2014;9:e99111.
- [4] Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing important health outcomes for comparative effectiveness research: an updated review and user survey. *PLoS One* 2016;11:e0146444.
- [5] Gorst SL, Gargon E, Clarke M, Smith V, Williamson PR. Choosing important health outcomes for comparative effectiveness research: an updated review and identification of gaps. *PLoS One* 2016;11:e0168403.
- [6] Davis K, Gorst SL, Harman N, Smith V, Gargon E, Altman DG, et al. Choosing important health outcomes for comparative effectiveness research: an updated systematic review and involvement of low and middle income countries. *PLoS One* 2018;13:e0190695.
- [7] Gargon E, Gorst SL, Harman NL, Smith V, Matvienko-Sikar K, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 4th annual update to a systematic review of core outcome sets for research. *PLoS One* 2018;13:e0209869.
- [8] Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: version 1.0. *Trials* 2017;18(Suppl 3):280.
- [9] Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core outcome set-standards for development: the COS-STAD recommendations. *PLoS Med* 2017;14(11):e1002447.
- [10] Sackett D, Richardson W, Rosenberg W, Haynes R. In: Evidence-based medicine: how to practice and teach EBM. New York: Churchill Livingstone.
- [11] Biggane AM, Brading L, Ravaud P, Young B, Williamson PR. Survey indicated that core outcome set development is increasingly including patients, being conducted internationally and using Delphi surveys. *Trials* 2018;19(1):113.
- [12] Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, et al. Core outcome set-standards for reporting: the COS-STAR statement. *PLoS Med* 2016;13(10):e1002148.
- [13] Gargon E, Crew R, Burnside G, Williamson PR. Higher number of items associated with significantly lower response rates in COS Delphi surveys. *J Clin Epidemiol* 2018;108:110–20.
- [14] MacLennan S, Kirkham J, Lam TBL, Williamson PR. A randomized trial comparing three Delphi feedback strategies found no evidence of a difference in a setting with high initial agreement. *J Clin Epidemiol* 2018;93:1–8.
- [15] Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *J Clin Epidemiol* 2018;96:84–92.